

Comprehensive Proteomic Profiling of the Membrane Constituents of a *Mycobacterium tuberculosis* Strain*

Sheng Gu‡, Jin Chen‡, Karen M. Dobos§, E. Morton Bradbury‡¶, John T. Belisle§, and Xian Chen‡||

***Mycobacterium tuberculosis* is an infectious microorganism that causes human tuberculosis. The cell membranes of pathogens are known to be rich in possible diagnostic and therapeutic protein targets. To compliment the *M. tuberculosis* genome, we have profiled the membrane protein fraction of the *M. tuberculosis* H37Rv strain using an analytical platform that couples one-dimensional SDS gels to a microcapillary liquid chromatography-nanospray-tandem mass spectrometer. As a result, 739 proteins have been identified by two or more distinct peptide sequences and have been characterized. Interestingly, ~450 proteins represent novel identifications, 79 of which are membrane proteins and more than 100 of which are membrane-associated proteins. The physicochemical properties of the identified proteins were studied in detail, and then biological functions were obtained by sorting them according to Sanger Institute gene function category. Many membrane proteins were found to be involved in the cell envelope, and those proteins with energy metabolic functions were also identified in this study. *Molecular & Cellular Proteomics* 2:1284–1296, 2003.**

Tuberculosis (TB)¹ is the major cause of death from an infectious disease in the world resulting in an estimated 8.5 million cases of clinical tuberculosis and 3 million deaths/year (1). The emergence of TB associated with HIV and multidrug-resistant TB has increased the threat to public health. The World Health Organization recognized the global emergency of TB in 1993. *Mycobacterium tuberculosis* (MTB), the etiologic agent of TB, can replicate in host cells by escaping host

cell defenses. The interactions between MTB and its host appear to be very delicately balanced (2, 3). Meanwhile, newly emerging drug-resistant strains of MTB are more difficult to cure, and they cause more fatal cases. Therefore, there are critical needs to identify new drugs or vaccines for MTB therapeutics. The recent completion of the genome sequence of the virulence MTB strain H37Rv (4) has provided new biomolecular insights into the mycobacterial cells.

Proteomics, the global analysis of the proteins expressed in a cell or tissue, provides a very promising approach for the large scale identification of proteins, their complexes, and their functions (5), which is required for the design of more effective and precise therapeutics or drug design (6). Therefore, the proteomic analysis of MTB strains is critical for an understanding of the molecular basis of its virulence and pathogenicity. A number of proteomic studies, mainly two-dimensional gel electrophoresis (2-DE)-based, have been carried out to identify proteins in various MTB strains and their subcellular localizations, including culture filtrate proteins and cell wall and cytosol fractions (7–12). These results demonstrate how the proteomics approaches complement genomics by profiling the protein products of the expressed genes. Further, the open reading frames in the MTB H37Rv strain that were not predicted from genomics were found by proteomics (13).

In the lipid environment of a cell membrane, many proteins are involved in important metabolic and biosynthetic processes (14). A functional genomics study of the pathogenicity of MTB implicated 16 genes in the virulence of this pathogen, and most of their products appeared to be involved in membrane transport and lipid metabolism (15). When MTB invades host cells, the capsule containing the mycobacterium forms just outside the membrane and cell wall of the bacterium. This interface between the host and the pathogen includes many important membrane surface enzymes/transporters involved in intercellular multiplication and the bacterial response to host microbicidal processes (16). Profiling the proteins or virulent factors actually expressed in the membrane compartment will reveal information on these pathways and possibly lead to the identification of new therapeutic targets. One study has focused on the very important membrane fraction of the mycobacterium to identify proteins responsible for MTB pathogenicity (17).

From the ‡Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, the §Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado 80523, and the ¶Department of Biological Chemistry, School of Medicine, University of California, Davis, California 95616

Received, June 30, 2003, and in revised form, October 3, 2003

Published, MCP Papers in Press, October 6, 2003, DOI 10.1074/mcp.M300060-MCP200

¹ The abbreviations used are: TB, tuberculosis; MTB, *Mycobacterium tuberculosis*; 1D, one-dimensional; 2-DE, two-dimensional gel electrophoresis; LC, liquid chromatography; μ LC, microcapillary LC; MS, mass spectrometry; MS/MS, tandem MS; MALDI, matrix-assisted laser desorption ionization; TOF, time-of-flight; TMH, transmembrane helix; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid.

From a genome-based prediction, in general, membrane proteins may represent ~30% of total gene products in a proteome (18). Because of the technical difficulties in obtaining samples of membrane proteins for physical characterization, less than 1% of the proteins have been identified as membrane proteins and are available in the Protein Data Bank. On the other hand, because of the good solubility of hydrophobic proteins in 1D SDS gels, Simpson *et al.* (19) demonstrated the use of an integrated approach involving 1D SDS gels and capillary LC-MS/MS for the identification of 92 membrane proteins in human cell lines. Our previous studies also showed that amino acid residue-specific mass-tagged membrane proteins could be solubilized and pre-separated in a 1D SDS gel and easily identified by MALDI MS (20). In this study, we have first obtained the membrane subcellular fraction of MTB through a differential centrifugation method. Separating this fraction on a 1D SDS gel to eliminate cellular components other than proteins, each gel slice was treated with trypsin, and the peptide digests were subjected to high resolution microcapillary LC separation prior to a highly sensitive nanospray-MS/MS analysis using a tandem quadrupole TOF tandem MS instrument. More than 700 proteins including some novel membrane and membrane-associated proteins were identified from this MTB membrane fraction. Very hydrophobic proteins, even those with 15 transmembrane helices, were also detected in this study. This comprehensive proteomic overview of the MTB membrane compartment revealed to some extent the molecular basis of MTB pathogenicity.

EXPERIMENTAL PROCEDURES

Growth of *M. tuberculosis* and Subcellular Fractionation—MTB H37Rv cells were cultured in 2 liters of glycerol alanine salts medium in roller bottles for 14 days at 37 °C with gentle agitation, washed with phosphate-buffered saline, pH 7.4, and inactivated by γ -irradiation. The culture supernatant and cells were separated by filtration through a 0.22- μ m membrane. 9 g of cells (total wet mass) were recovered from the culture and resuspended in 4.5 ml of TSE (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA) supplemented with a protease inhibitor mixture (0.7 μ g/ml pepstatin A, 0.5 μ g/ml leupeptin, 0.2 mM phenylmethylsulfonyl fluoride). DNase and RNase (0.6 μ g/ml) were added, and the cells were broken by sequential passage through a French press cell (Thermo IEC, Needham, MA). Lysis of the tubercle bacilli was confirmed by acid fast staining. The suspension was centrifuged at 2500 \times g to remove any unbroken cells and debris. The clarified lysate was centrifuged at 27,000 \times g at 4 °C for 1 h. The supernatant was then centrifuged at 100,000 \times g at 4 °C for 4 h, and the pellet was recovered as cell membrane. The membrane pellets were washed with phosphate-buffered saline followed by centrifugation at 27,000 \times g for 1 h and at 100,000 \times g for 4 h. The membrane pellets were resuspended into 0.1 M NH₄HCO₃ by bath sonication until a uniform suspension was obtained and were dialyzed against 0.1 M NH₄HCO₃, 1 mM dithiothreitol, followed by dialysis against 0.1 M NH₄HCO₃. Approximately 13 mg of cell membrane proteins were recovered.

One- and Two-dimensional Gel Electrophoresis—In the 1D gel approach, 200 μ g of protein were mixed with 50 μ l of SDS loading buffer and boiled for 10 min before loading on a 10-cm-long, 1-mm-thick 12% SDS-polyacrylamide gel. The protein bands were visualized with Coomassie Brilliant Blue R-250 staining (Bio-Rad). The gel

bands were cut continuously in the range of 8 kDa up to the loading well. A total of 45 bands were excised from the gel, and the gel slices were placed into 1-mm³ cubes for further destaining and digestion.

For a 2-DE gel, 100 μ g of protein were added to 350 μ l of rehydration buffer (7 M urea, 2 M thiourea, 4% CHAPS, 2% dithiothreitol, 2% IPG buffer, pH 4–7, 1 mM benzamide, 1.5 mM EDTA/EGTA, 1 mM sodium vanadate, 1 μ M microcystin-LR, 2 μ g/ml pepstatin-A, 10 μ g/ml aprotinin, 20 μ g/ml leupeptin) and rehydrated at 20 V for 12 h. Isoelectric focusing was performed at ambient temperature on an 18-cm Immobiline pH 4–7 dry strip using a IPGphor II isoelectric focusing system (Amersham Biosciences). The voltage was held at 500 V for 30 min and then stepped to 1000 V and held for 2 h; next a 3-h linear gradient from 1000 to 7000 V was applied to the strips. The voltage was held at 7000 V for another 4 h. Prior to the second dimension, the strips were incubated for 10 min in equilibration buffer (6 M urea, 2% SDS, 0.375 M Tris, pH 8.8, 20% glycerol) first with 130 mM dithiothreitol and second with 135 mM iodoacetamide. Each equilibrated strip was then put onto a 20 \times 20-cm 10% Duracryl gel, and the second dimension was run at 500 V for 5 h in the two-dimensional gel system (Genomic Solutions, Ann Arbor, MI). The gel was silver-stained as described previously (21). The protein spots and bands were excised from the gel and in-gel digested as described previously (22).

High Performance Microcapillary Liquid Chromatography-Tandem Mass Spectrometry (μ LC-MS/MS)—The in-gel digest of each gel slice was analyzed by a μ LC-nanospray-MS/MS using a QSTAR Pulsar I mass spectrometer (Applied Biosystems, Foster City, CA) coupled with LC Packings Ultimate microcapillary LC system (Dionex, Sunnyvale, CA). The PepMap C₁₈ column (3 μ m, 100 Å, 75- μ m inner diameter, 15-cm length) employed for peptide separation was also purchased from Dionex. The autosampler was configured using the partial loop injection mode involving a 10- μ l sampling loop; a preconcentration C₁₈ cartridge was connected with the analytical column through a 10-port switch valve. The partial loop injection method loaded a 3- μ l sample into the 10- μ l loop, which was then pumped onto the preconcentration column at a flow rate of 30 μ l/min by a sample-loading pump. Three min after the start of the sample loading, the 10-port valve was used to switch the preconcentration cartridge in line with the nanoflow solvent delivery system, thus enabling the trapped peptides to be eluted onto the analytical column. Mobile phase A is 0.1% formic acid and 5% acetonitrile. Mobile phase B is 0.1% formic acid and 95% acetonitrile. The gradient was kept at 5% B for 5 min, then ramped linearly from 5 to 50% B in 50 min, and then jumped to 75% B and kept for 10 min. Then the gradient was jumped back to the start point, and the column was equilibrated for 10 min. The flow rate was 200 nl/min.

The end of the analytical column was connected with a 10- μ m inner diameter PicoTip nanospray emitter (New Objective, Woburn, MA) by a stainless steel union (Valco Instrument, Houston, TX) mounted on the nanospray source (Protana Engineering, Odense, Denmark). The spray voltage (usually set between 1800 and 2100 V) was applied to the emitter through the stainless steel union and tuned to get the best signal intensity using standard peptides. The two most intense ions with charge states between 2 and 4 in each survey scan were selected for the MS/MS experiment, provided they passed the switching criteria of MS/MS scan. The rolling collision energy feature was employed to fragment the peptide ions according to their charge states and *m/z* values.

MALDI-TOF MS Peptide Mass Mapping—All MALDI-TOF mass spectra were acquired with a PE Voyager DE_STR biospectrometry work station equipped with an N₂ laser (337 nm, 3-ns pulse width, 20-Hz repetition rate) using the reflector mode with delayed extraction (Applied Biosystems). The matrix, α -cyano-4-hydroxycinnamic acid, was prepared as a saturated solution in 50% acetonitrile, 0.1% trif-

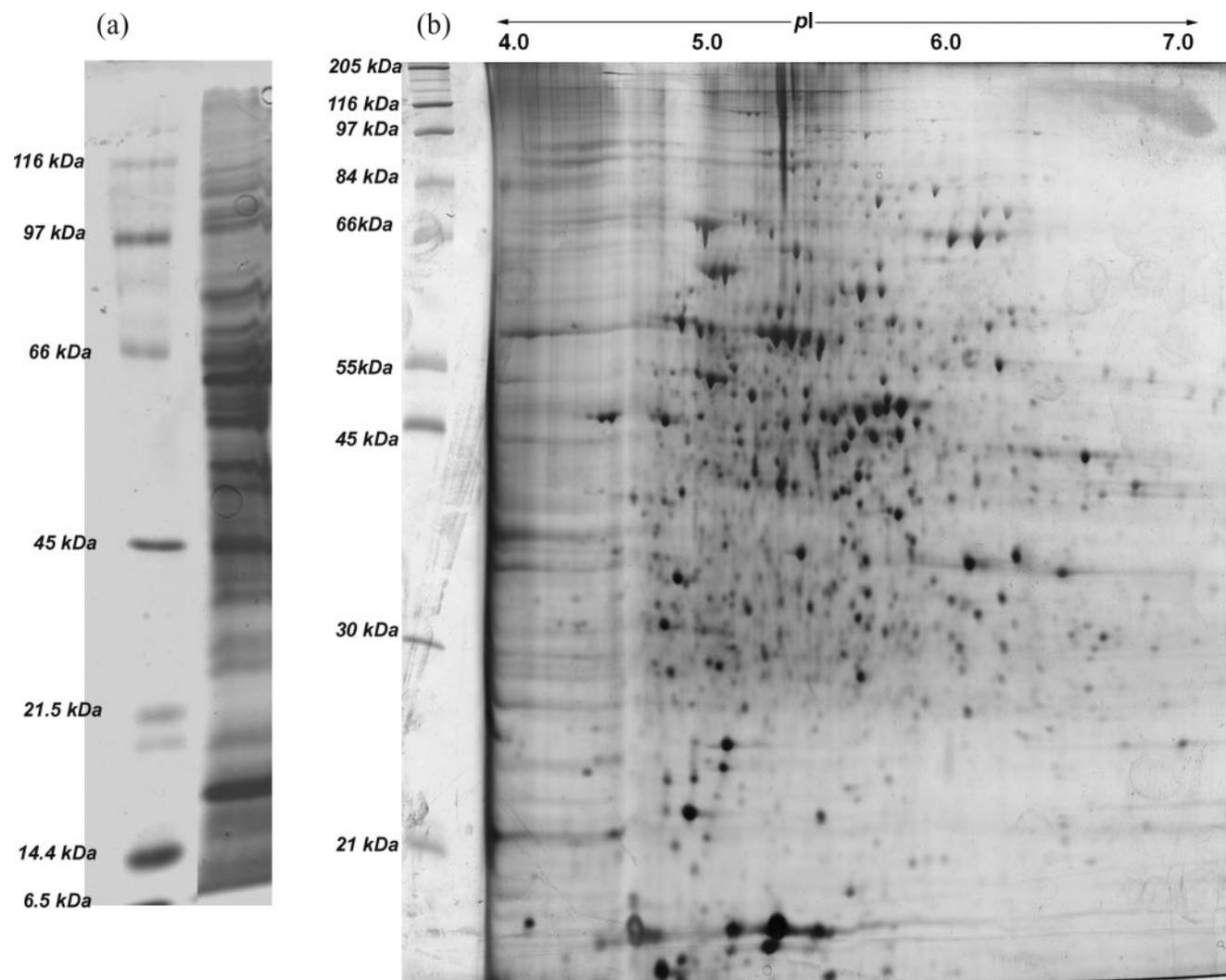


FIG. 1. Gel images of protein separation of the *M. tuberculosis* membrane fraction using (a) 12% SDS-PAGE and (b) a 2-DE gel in the range of pH 4–7 with 10% Duracryl.

luoroacetic acid solvent. For MALDI-TOF analysis, 0.5 μ l of the matrix solution was mixed with 0.5 μ l of sample on the sample plate, and the mixture was air-dried to form the crystal analyte.

Protein Database Searching—Each MALDI-TOF peptide mass mapping was first calibrated with the two standard peptides (angiotensin II and oxidized insulin chain B) and then submitted to the database (NCBI nr.10.30.2002) to determine the protein identities using the MS-Fit program (University of California-San Francisco Mass Spectrometry Facility). The ProID® program loaded on the QSTAR instrument was used to interpret the LC-MS/MS data by searching against the MTB H37Rv strain databases from the National Center for Biotechnology Information (NCBI). Post-translational modifications including carboxymethylation of cysteine, methionine oxidation, and phosphorylations on serine, tryptophan, and tyrosine were also considered as possible modifications in the database search.

RESULTS

Comprehensive Protein Profiling of the MTB Membrane Fraction Using the Integrated SDS Gel- μ LC-MS/MS Approach—In our study, 1D SDS gel separation of the membrane fraction was performed prior to LC-MS/MS analysis.

For the purpose of the comparison, the same fraction was resolved by the commonly used approach of 2-DE in the similar molecular mass range as shown in Fig. 1. We noted that there was a clear similarity in the 2-DE spot distribution pattern obtained from the membrane and cytosol fractions, suggesting certain highly abundant soluble proteins co-existing in the membrane fraction. The better solubility of those cytosolic proteins in a 2-DE may overwhelm low abundance membrane proteins from being visualized and analyzed in mass spectrometry. 2-DE gel separation coupled with MS identified a limited number of proteins from the more than 600 silver-stained spots on each gel. For example, the acquired MALDI-TOF MS data (100 spectra from 64 excised 2-DE protein spots) led to the identification of 33 proteins as listed in Table I. With pH 4–7 immobilized pH gradient strips employed, our identification results covered those proteins with theoretical pIs from 4.50 to 6.30 and molecular masses from 13.4 to 70.2 kDa. Two membrane proteins with one predicted

TABLE I
A list of proteins identified based on 2-DE coupled with MALDI-TOF MS

| RefSeq ID | Sanger ID | SwissProt ID | Gene name | Description | Molecular mass (kDa)/pI ^a | Reported previously ^b |
|-------------|-----------|--------------|---------------|--|--------------------------------------|----------------------------------|
| gi 15607523 | Rv0382c | O53717 | <i>umpA</i> | probable uridine 5'-monophosphate synthase | 18.9/5.23 | |
| gi 15607792 | Rv0652 | P37381 | <i>rplL</i> | 50 S ribosomal protein L7/L12 | 13.4/4.59 | 7, 8, 9, 10 |
| gi 15607817 | Rv0677c | O53785 | <i>mmpS5</i> | conserved small membrane protein | 16.1/5.00 | |
| gi 15607825 | Rv0685 | P31501 | <i>tuf</i> | elongation factor EF-Tu | 43.6/5.28 | 7, 8, 9, 17 |
| gi 15607873 | Rv0733 | O53796 | <i>adk</i> | probable adenylate kinase | 20.2/5.02 | 7, 8, 9, 10 |
| gi 15607964 | Rv0824c | Q50824 | <i>desA1</i> | acyl-[ACP] desaturase | 38.7/6.21 | 7, 9, 13 |
| gi 15608045 | Rv0905 | Q10533 | <i>echA6</i> | enoyl-CoA hydratase/isomerase superfamily | 26.6/5.85 | 7, 9, 13 |
| gi 15608086 | Rv0946c | P77895 | <i>pgi</i> | glucose-6-phosphate isomerase | 60.0/5.40 | |
| gi 15608163 | Rv1023 | P96377 | <i>eno</i> | enolase | 44.9/4.50 | 17 |
| gi 15608450 | Rv1310 | Q10593 | <i>atpD</i> | ATP synthase β chain | 53.1/4.86 | 17 |
| gi 15608469 | Rv1329c | Q10640 | <i>dinG</i> | probable ATP-dependent helicase | 70.2/6.10 | |
| gi 15608531 | Rv1392 | P77899 | <i>metK</i> | S-adenosylmethionine synthase | 43.0/4.95 | 7, 9, 11 |
| gi 15608684 | Rv1546 | Q10780 | | hypothetical protein Rv1546 | 15.3/6.30 | |
| gi 15608690 | Rv1552 | Q10760 | <i>frdA</i> | fumarate reductase flavoprotein subunit | 63.7/5.88 | |
| gi 15608796 | Rv1658 | P94993 | <i>argG</i> | arginosuccinate synthase | 43.7/5.26 | |
| gi 15609168 | Rv2031c | P30223 | <i>hspX</i> | 14-kDa antigen, heat shock protein Hsp20 family | 16.1/5.00 | 7, 8, 9, 10, 12, 17 |
| gi 15609234 | Rv2097c | Q10706 | | hypothetical protein Rv2097c | 51.4/6.26 | |
| gi 15609382 | Rv2245 | Q10524 | <i>kasA</i> | 3-oxoacyl-[acyl-carrier protein] synthase | 43.3/5.11 | |
| gi 15609357 | Rv2220 | Q10377 | <i>glnA1</i> | glutamine synthase class I | 53.6/5.04 | 8, 9 |
| gi 15609383 | Rv2246 | Q10525 | <i>kasB</i> | β -ketoacyl-ACP synthase (meromycolate) | 44.2/5.27 | 7, 9 |
| gi 15609513 | Rv2376c | O05822 | | hypothetical protein Rv2376c | 15.1/6.05 | 8, 9 |
| gi 15609597 | Rv2460c | O53187 | <i>clpP2</i> | ATP-dependent Clp protease proteolytic subunit 2 | 23.5/4.99 | |
| gi 15609598 | Rv2461c | O53188 | <i>clpP</i> | ATP-dependent Clp protease proteolytic subunit 1 | 21.7/4.79 | 10, 17 |
| gi 15609648 | Rv2511 | O06174 | | hypothetical protein Rv2511 | 23.4/4.53 | |
| gi 15609726 | Rv2589 | Q50632 | <i>gabT</i> | 4-aminobutyrate aminotransferase | 46.7/5.38 | |
| gi 15609917 | Rv2780 | P30234 | <i>ald</i> | L-alanine dehydrogenase | 38.7/5.81 | 7, 8, 9, 11 |
| gi 15610019 | Rv2882c | Q10794 | <i>frr</i> | ribosome recycling factor | 20.8/5.71 | 7, 8, 9 |
| gi 15610138 | Rv3001c | O53248 | <i>ilvC</i> | ketol-acid reductoisomerase | 36.1/5.03 | |
| gi 15610166 | Rv3029c | O53276 | <i>fixA</i> | electron transfer flavoprotein β subunit | 28.1/4.66 | 7, 8, 9, 10 |
| gi 15610410 | Rv3274c | P96879 | <i>fadE25</i> | acyl-CoA dehydrogenase | 41.7/5.21 | 7, 9 |
| gi 15610416 | Rv3280 | P96885 | <i>accD5</i> | acetyl/propionyl CoA carboxylase β subunit | 59.4/5.19 | 7, 9 |
| gi 15610475 | Rv3339c | O53389 | <i>icd1</i> | isocitrate dehydrogenase | 45.5/5.13 | |
| gi 15610593 | Rv3457c | O06324 | <i>rpoA</i> | α subunit of RNA polymerase | 37.7/4.64 | 7, 8, 9 |
| gi 15610989 | Rv3853 | P96224 | <i>menG</i> | S-adenosylmethionine:2-demethylmenaquinone | 16.2/5.00 | |

^a Molecular mass and pI were calculated by the M_r/pI tools at the ExPASy server.

^b Listed are references in which the proteins have been identified previously by proteomic analysis.

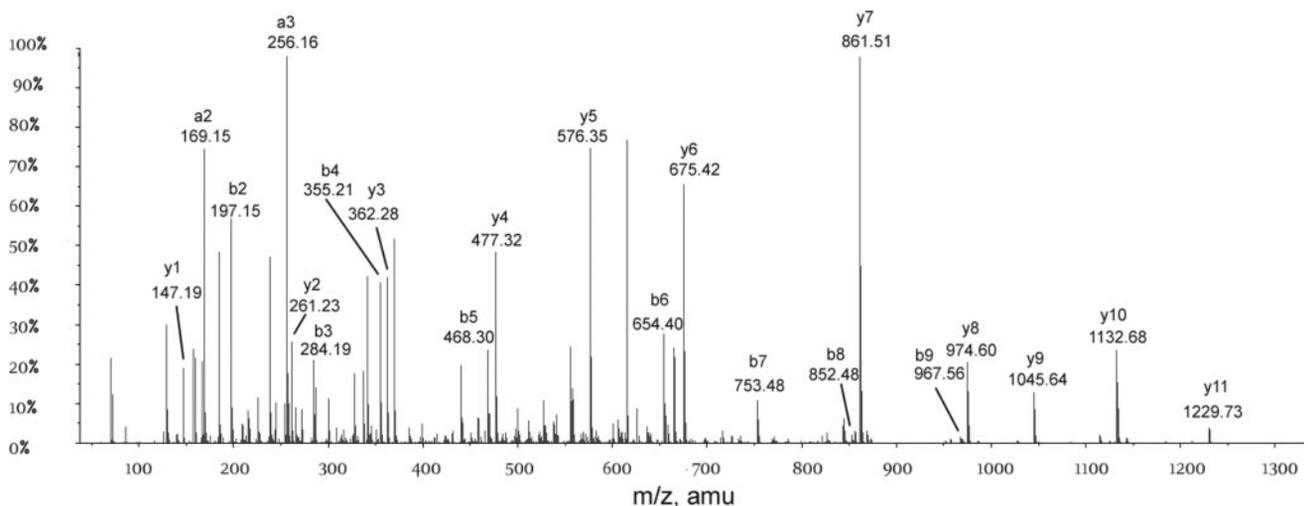


FIG. 2. A tandem MS spectrum of a peptide at 1327.81 Da. This spectrum matched to the peptide sequence VPSAIWVVDTNK with a match score of 30 in the database searching, leading to the identification of its parent protein, rpsB (gi|15610027). Another five peptides from the same protein were also identified with scores higher than 30. *amu*, atomic mass unit.

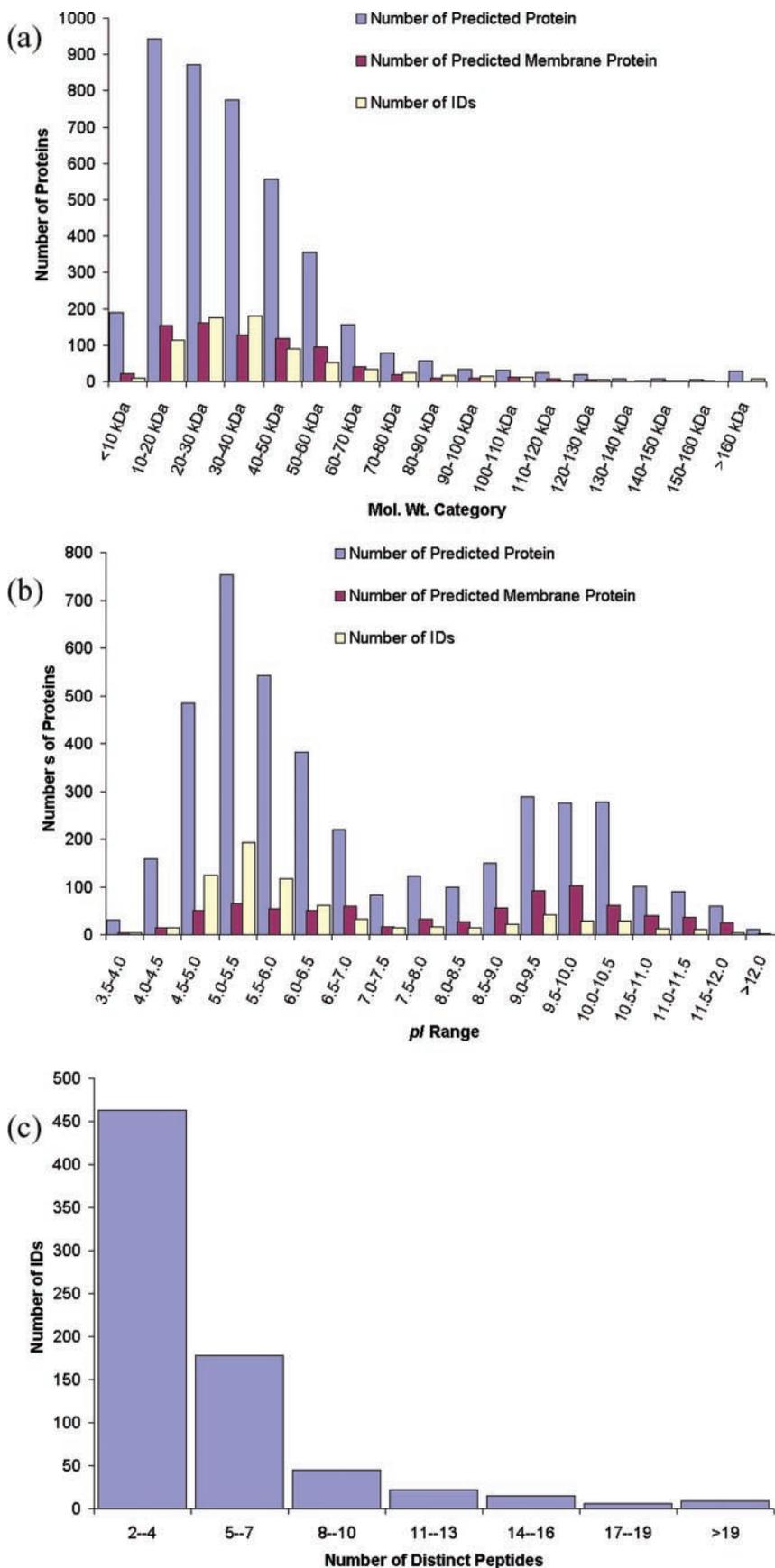
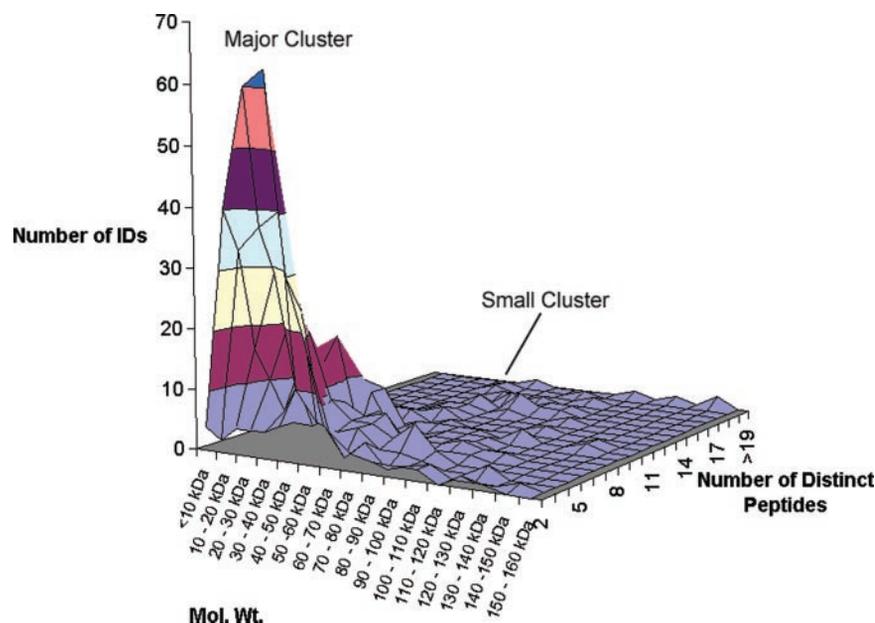


FIG. 3. The distribution of molecular mass and pI of the total predicted proteins, the predicted membrane proteins, and the actually identified proteins. a, molecular mass distribution with 10-kDa molecular mass increments. b, pI distribution of the proteins with 0.5 pH unit increments. c, distribution of the number of distinct peptides for the proteins identified. ID, identification.

FIG. 4. A profile of the number of proteins identified based on both molecular mass and the number of distinct peptides matched in the database searching. In addition to the *major cluster* at 10–50 kDa and 2–5 distinct peptides identified, there was a *small cluster* at 80–120 kDa (Mol. Wt.) and 7–12 distinct peptides identified. ID, identification.



transmembrane helix (Rv0677c, Rv2367c) were identified by this 2-DE-based method. Another three proteins (Rv1310, Rv2031c, and Rv2461c) were reported previously as membrane-associated proteins by Sinha and co-workers (17).

A total of 45 consecutive 1- to 2-mm gel slices were obtained from the 1D SDS gel across the entire gel lane. The in-gel digestion extracts were vacuum-dried and resuspended in 5% acetonitrile, 95% water, 0.1% trifluoroacetic acid solution. No enrichment or desalting procedures were performed prior to sample submission for on-line μ LC high resolution separation and MS/MS peptide sequencing. Each MS/MS spectrum was searched against the database using the ProID[®] search engine. The searching criteria set for any acceptable match were better than 98% confidence and a peptide sequence matching score greater than 30, which was considered as a high score that gave unambiguous protein identification. An example of an MS/MS spectrum with a match score of 30 is shown in Fig. 2. The MS/MS fragment ion pattern precisely matched to the peptide sequence of the parent protein rpsB (gi|15610027). In the same LC-MS/MS run, another five peptides that originated from this protein were also identified with scores better than 30. Individual proteins matched with two or more peptide sequences were considered to be unambiguous identifications. These identifications were further validated by searching against a reversed sequence TB H37Rv database using Gygi's method to evaluate the multidimensional LC-MS/MS data (23). No match exceeded the criteria described above. Applying the above identification criteria to these 45 gel slices, a total of 739 proteins were identified in the MTB membrane fraction (Supplementary Table I). As seen in our dataset, many hypothetical proteins were found to be expressed in the MTB H37Rv strain. There were some MS/MS spectra that did not result in any identifications of MTB proteins. Those MS/MS spectra gave

out ambiguous database search results and therefore did not pass the threshold for positive identification. Some unmatched signals were identified as having originated from trypsin autolysis and human keratins.

Our protein identification dataset gave a much wider range of proteins in the membrane fraction than those identified from the 2-DE-based method. We first categorized total predicted proteins, predicted membrane proteins, and identified proteins according to their different physical properties. For example, in the molecular mass distribution shown in Fig. 3a, the proteins in the 20–30- and 30–40-kDa molecular mass ranges are the dominant species in the membrane compartment. Note also a small number of large proteins (>160 kDa). In the MTB proteome, the most of the gene products are predicted in the 10–20-kDa molecular mass range. However, the number of proteins actually identified in the 10–20-kDa molecular mass range was not the maximum number in the molecular mass category. The proteins in a wide pI range from 3.5 to 12 were also simultaneously identified; the protein with the maximum pI at 11.95 was the DNA-binding protein II (Rv2986c). On the pI distribution plot (Fig. 3b), there were also two abundant clusters in the pI ranges of 4.5–7.0 and 8.5–10.5. This distribution pattern is very similar to that predicted previously using a computational approach (24). As seen in the distribution pattern, the cluster around pI 5.0 has many more proteins than in the second cluster. In this study, the second cluster (pI 8.5–10.5) was found to contain a larger number of membrane proteins than for any other pI ranges. The pI distribution of the total predicted proteins, the predicted membrane proteins, and the actually identified proteins had the same distribution pattern.

The majority, ~60%, of the identified proteins had 2–4 distinct peptide sequences matched to each protein. 7% of proteins were identified through the matches of 10 or more

TABLE II
Membrane proteins identified by 1D SDS-PAGE coupled with microcapillary LC-MS/MS

| RefSeq ID ^a | No. of TMHs ^b | Topology ^c |
|------------------------|--------------------------|---|
| gi 15607150 | 1 | i102-124o |
| gi 15607161 | 1 | o6-28i |
| gi 15607190 | 1 | i89-111o |
| gi 15607214 | 4 | i16-38o231-253i274-296o306-328i |
| gi 15607286 | 1 | o235-257i |
| gi 15607299 | 9 | o5-27i34-53o58-77i90-112o127-149i170-192o197-219i226-245o249-268i |
| gi 15607308 | 6 | o15-37i56-78o93-115i155-177o197-219i240-262o |
| gi 15607311 | 1 | i7-29o |
| gi 15607317 | 3 | i29-51o56-78i166-188o |
| gi 15607318 | 1 | o26-48i |
| gi 15607319 | 1 | i79-101o |
| gi 15607321 | 6 | i25-47o218-240i269-291o301-323i330-352o392-414i |
| gi 15607341 | 3 | i7-29o33-55i75-97o |
| gi 15607368 | 2 | i5-27o319-341i |
| gi 15607424 | 1 | i72-94o |
| gi 15607425 | 1 | i45-67o |
| gi 15607431 | 11 | i129-151o155-177i184-202o206-228i235-254o258-280i328-347o352-374i381-400o404-426i447-469o |
| gi 15607432 | 1 | o431-453i |
| gi 15607433 | 2 | i13-30o35-57i |
| gi 15607479 | 6 | o5-24i71-93o113-135i148-170o180-202i214-231o |
| gi 15607483 | 1 | o497-519i |
| gi 15607524 | 1 | o5-24i |
| gi 15607533 | 1 | o387-409i |
| gi 15607559 | 1 | i7-29o |
| gi 15607620 | 1 | i113-135o |
| gi 15607646 | 1 | o345-367i |
| gi 15607696 | 2 | i33-55o131-153i |
| gi 15607826 | 2 | i9-31o35-57i |
| gi 15608015 | 1 | i5-27o |
| gi 15608041 | 1 | o4-26i |
| gi 15608071 | 1 | o380-402i |
| gi 15608109 | 6 | i141-163o216-238i375-397o402-424i718-737o741-760i |
| gi 15608125 | 2 | i13-32o69-91i |
| gi 15608218 | 2 | i95-117o199-221i |
| gi 15608374 | 2 | i85-107o112-134i |
| gi 15608410 | 1 | i7-24o |
| gi 15608444 | 5 | o30-47i85-107o127-149i189-211o221-240i |
| gi 15608446 | 1 | o32-54i |
| gi 15608447 | 1 | i5-23o |
| gi 15608488 | 6 | o293-315i336-358o407-429i434-456o520-542i549-571o |
| gi 15608502 | 1 | i60-82o |
| gi 15608508 | 1 | i20-37o |
| gi 15608555 | 2 | i20-39o54-71i |
| gi 15608614 | 1 | o138-160i |
| gi 15608619 | 3 | o15-37i67-86o307-329i |
| gi 15608626 | 1 | i7-29o |
| gi 15608729 | 4 | i25-47o74-96i108-130o167-189i |
| gi 15608920 | 1 | i54-76o |
| gi 15608932 | 10 | i136-158o168-190i203-220o224-246i251-270o274-296i357-379o411-433i440-462o477-499i |
| gi 15609023 | 1 | i20-42o |
| gi 15609114 | 1 | i155-177o |
| gi 15609129 | 5 | i160-182o204-226i378-400o410-432i710-732o |
| gi 15609188 | 7 | i21-39o43-60i67-89o99-121i128-150o165-187i206-228o |
| gi 15609332 | 3 | i97-119o134-156i202-224o |
| gi 15609333 | 9 | o46-68i115-134o144-166i178-200o215-237i267-289o333-355i375-397o412-434i |
| gi 15609337 | 3 | i31-53o73-95i116-138o |
| gi 15609351 | 1 | i164-186o |
| gi 15609361 | 1 | i13-35o |
| gi 15609421 | 3 | i7-29o44-66i78-95o |

TABLE II—continued
Membrane proteins identified by 1D SDS-PAGE coupled with microcapillary LC-MS/MS

| RefSeq ID ^a | No. of TMHs ^b | Topology ^c |
|------------------------|--------------------------|--|
| gi 15609426 | 1 | i7-29o |
| gi 15609462 | 4 | i19-36o51-73i86-108o123-145i |
| gi 15609463 | 4 | o46-68i81-103o129-151i195-217o |
| gi 15609482 | 3 | i7-26o162-181i609-631o |
| gi 15609700 | 4 | i16-38o231-253i274-296o306-328i |
| gi 15609719 | 1 | i47-69o |
| gi 15609723 | 6 | i66-88o191-208i215-237o241-263i301-323o328-350i |
| gi 15609725 | 1 | o4-21i |
| gi 15609734 | 1 | i2-24o |
| gi 15609762 | 6 | i13-35o45-64i77-99o109-131i138-160o207-229i |
| gi 15610058 | 1 | i5-27o |
| gi 15610082 | 1 | i7-24o |
| gi 15610106 | 1 | o20-42i |
| gi 15610227 | 2 | o5-24i37-59o |
| gi 15610238 | 4 | o20-42i174-196o225-247i268-290o |
| gi 15610329 | 7 | i17-39o59-81i113-135o169-191i212-229o255-277i284-306o |
| gi 15610409 | 10 | o36-58i60-82o102-124i129-151o205-227i264-286o306-328i341-360o364-386i393-424o |
| gi 15610526 | 1 | i13-32o |
| gi 15610573 | 1 | o73-95i |
| gi 15610720 | 1 | i13-35o |
| gi 15610746 | 2 | o10-27i111-133o |
| gi 15610818 | 1 | i7-29o |
| gi 15610819 | 1 | i7-29o |
| gi 15610940 | 1 | i19-41o |
| gi 15610942 | 7 | o55-72i92-114o118-140i161-183o216-238i245-264o284-301i |
| gi 15611046 | 15 | i40-62o66-88i105-127o142-164i176-198o213-235i255-276o291-313i343-365o375-397i410-432o437-459i480-497o517-539i979-1001o |

^a Shown are RefSeq identifications from the National Center for Biotechnology Information; novel identifications are shown in bold.

^b The number of transmembrane helices predicted by TMHMM 2.0.

^c The topology predicted by TMHMM 2.0 is given as the amino acid position (number) of the transmembrane helices from the N terminus of the protein. “i” indicates that the loop is on the inside, and “o” indicates that the loop is on the outside.

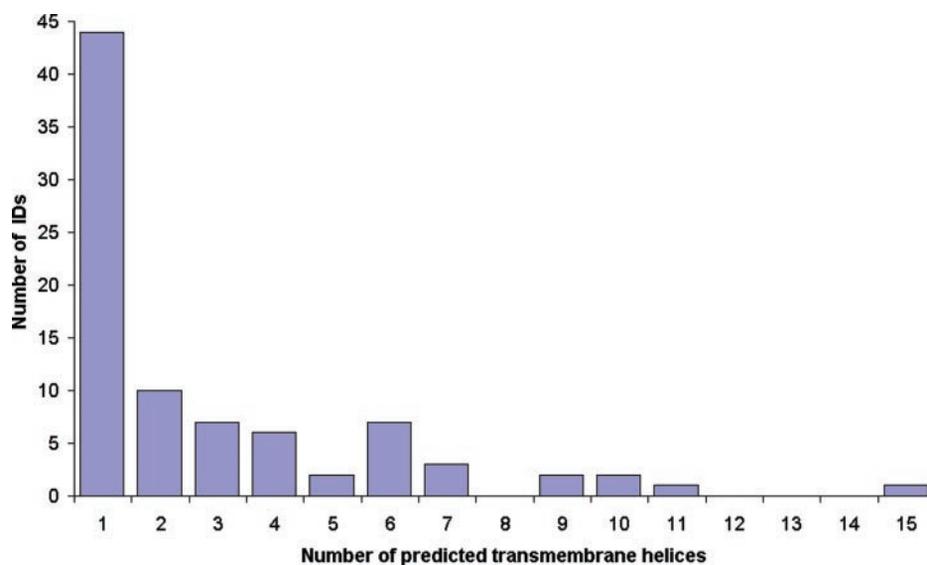
unique peptide sequences for each protein (as shown in Fig. 3c). The maximum number of 44 unique peptide sequences was found to match to a single protein identified as DNA-directed RNA polymerase β -chain (Rv0668) with a molecular mass of 146,769 Da. Fig. 4 shows a three-dimensional map of the correlations among the molecular mass, the number of distinct peptides, and the number of proteins identified. In the global picture of our identified protein profile, the majority of the proteins in the 10–50-kDa molecular mass range have been identified with 2–5 characteristic peptides. Proteins found in a small cluster in the molecular mass range of 80–120 kDa were identified by matching 7–12 distinct peptide sequences for each of them. It also suggests that there were some fairly abundant proteins between 80 and 120 kDa in the MTB membrane fraction.

A Study of Topology and Transmembrane Domain Localization of the Proteins Profiled in the Membrane Fraction—Several of the computational methods used in membrane protein topology analysis have provided the basis for the characterization of the MTB membrane fraction. Many tools are available to predict the topology of membrane proteins, and their prediction accuracy has been improved by various

modeling systems (25, 26). TMHMM 2.0 (27, 28), one of the best of the performance programs (26), was used to identify transmembrane proteins in our data profile. The TMHMM 2.0 program predicts protein topology based on the sequences in FASTA format. Of the 739 identified protein sequences submitted for characterization, TMHMM 2.0 has identified 85 transmembrane proteins as summarized in Table II. Fig. 5 shows the distribution of transmembrane helix (TMH) numbers for our identified membrane proteins. More than half of the transmembrane proteins have one TMH. Other transmembrane proteins have TMHs ranging from 2 to 15, although none of them have 8, 12, 13, or 14 TMHs. Notably, the hypothetical protein Rv3910 with 15 TMHs was identified through the presence of four distinct peptide sequences.

From a genome-wide topology prediction using TMHMM 2.0, there are 787 possible transmembrane proteins with 1 to 16 or 18 TMHs, which is consistent with the genomic data provided by PEDANT (pedant.gsf.de). 10.8% of these membrane proteins were identified in our study, whereas for the proteins with 1 TMH the identification rate was 15.6%. There are 66 membrane proteins with an even number of 6 TMHs existing in the MTB proteome, much higher than those with 5 or 7 TMHs. This

FIG. 5. **Distribution of transmembrane helices of the MTB membrane proteins identified.** TMHMM 2.0 software was used to predict the number of TMHs in each of the identified MTB protein. The proteins with one predicted TMH comprise approximately 50% of the total number of proteins with TMHs. The number of identified proteins is proportional to the total protein number at each different TMH number.



trend is also true for all identified membrane proteins in our profile. Thus, the success of our strategy is much improved over the average proteomic analysis. This approach of 1D SDS gel-LC-MS/MS is proven to be fully capable of the identification and characterization of hydrophobic proteins.

The membrane proteins (Rv0969, Rv3273) were chosen to demonstrate the use of sequence and topology in their characterization. ctpV (Rv0969) is known as an integral membrane protein involved in ATP/ADP catalysis and cation transportation. It is also transcribed during the response of the host cell to MTB infection (29). This protein contains 6 TMHs as indicated in Fig. 6a. Six non-redundant peptides represent 14.9% of the protein sequence. In addition, these peptides are found in the sequences at both the inner and outer membrane segments. Rv3273, which has 10 TMHs, is also an integral membrane protein similar to carbonic anhydrase. 11.1% of this protein sequence was identified in our study. Interestingly, a peptide covering the seventh TMH from the N terminus was also sequenced by MS/MS (Fig. 6b), suggesting that the transmembrane segments of the proteins in the membrane compartment are accessible to proteolytic enzymes.

Categorization of Functional Groups in the Proteomic Profile—The predicted proteins in the MTB H37Rv strain can be categorized in six major functional groups and 32 subgroups based on the annotation developed at the Sanger Institute (4). The classification of the 739 identified proteins in these main and sub-functional groups is shown in Fig. 7, a and b. In Fig. 7a, 38.3% of the identified proteins are involved in small molecule metabolism, which is further subdivided according to the functions in Group I of Fig. 7b. Similarly, 21.7% of the identified proteins are involved in the metabolism of macromolecules, and these proteins are further subdivided according to their functions in Group II of Fig. 7b. Therefore, 60% of the 739 identified proteins are involved in various metabolic pathways. Another 6.6% of the protein profile (Fig. 7a) is

required for various cell processes as further subdivided in Group III of Fig. 7b. The proteins classified into the functional Groups IV (other), V (conserved hypothetical), and VI (unknown functions) only represented 33.3% of our profile. In the protein profile of the functional Group I (Fig. 7b), 32% of the proteins are involved in energy metabolism. Among the proteins in functional Group II (Fig. 7b), 52% are involved in synthesis and modification of macromolecules, and 35% are in the cell envelope. We have also compared the number of the identified proteins with the total open reading frames predicted in each sub-functional group (Fig. 7c). Our proteomic profile represents 25% of the total predicted gene products in the major Groups I, II, and III. Also, only 4.2% of the total predicted gene products were found in major functional Group IV, suggesting fewer proteins in functional Group IV expressed in the membrane fraction.

The 85 identified membrane proteins were distributed in all of the six major functional groups; 14, 35, 7, 3, 18, and 8 were found in the Groups I, II, III, IV, V, and VI, respectively. Notably, 32 of the 35 membrane proteins in Group II were in the subgroup II.C (cell envelope). In this subgroup, there are lipoproteins, surface polysaccharides, proteins, and antigens, and other membrane proteins. These proteins contain many hydrophobic sequences probably required to locate them in the cell membrane. Some of the identified proteins are transporter and channel proteins, such as Rv0985c and Rv0072, and some are secreted proteases, such as Rv0291.

Also, nine of the total 16 chaperone/heat-shock proteins were identified in the MTB membrane fraction. Most of them were identified unambiguously through good sequence coverage with an average of 7.6 peptide sequences matched for each protein. This clearly shows that these trafficking proteins are extensively involved in transportation through the cell membrane, thus remaining highly abundant in the membrane fraction after subcellular fractionation.

(a)
 gi|15608109 Probable cation-transporting ATPase V
 MRVCVTGFNV DAVRAVAIEE TVSQVTGVHA VHAYPRTASV VIWYSPELGD
 TAAVLSAITK AQHVPAELVP ARAPHSAGVR GVGVVVRKITG GIRRMLSRPP
 GVDKPLKASR CGRRPRGPVR GSASWPGEQN RRERRTWLPR **VWLALPLGLL**
ALGSSMFFGA YPWAGWLAF A TLPPVQFVAG WPILRGAVQQ ARALTSNMDT
 LIALGTLTAF VYSTYQLFAG **GPLFEDTSAL** IIAFVVLGRH LEARATGKAS
 EAISKLELGG AKEATLLVDG **OELLVPVDQV** QVGDVLRVVRP GEKIPVDGEV
 TDGRAAVDES **MLTGESVPVE** KTAGDRVAGA TVNLDGLLTV RATAVGDATA
 LAQIVRLVEQ AQGDKAPVQR LADRVSAVAVF **PAVIGVAVAT** FAGWTLLAAN
 PVAGMTAAVA VLIACPCAL **GLA** TPTAIMV GTGRGAELGI LVKGGVEVLEA
 SKKIDTVVFD KTGLTRARM RVTDVIAQQR RQPDQVLRLA **AAVESGSEHP**
IGAAIVAAAH ERGLAIPAAN AFTAVAGHGV RAQVNGGPPV VGRKRLVDEQ
 HLVLDPHLAA AAVEQEERGR TAVFVGDGQ VVGVLAVADT VKDDAADVVG
 RLHAMGLQVA MITGDNARTA AAIAKQVIE KVLAEVLPQD **KVAEVRRLQD**
 QGRVVAMVGD GVNDAPALVQ ADLGLAIGTG TDVAIEASDI TLMGRLDGV
 VRAIELSRQT LRTIYQNLGW AFGYNTAAIP **LAALGALNPV** VAGAAMGFSS
VSVVTNSLR RRFGRDGRTA

(b)
 gi|15610409 hypothetical protein Rv3273
 MTIPRSQHMS TAVNSCTEAP ASRSQWMLAN LRHDVPASLV VFLVALPLSL
 GIAIASGAP IAGVIAAVVG GIVAGAVGGS PVQVSGPAAG LTVVVAELID
 ELGWPMILCLM TIAAGALQIV FGLSRMARAA LAIAPVVVHA MLAGIGITIA
 LQQIHVLLGG TSHSSAWRNI VALPDGILHH ELHEVIVGGT VIAILLMWSK
 LPAKVRIIPG PLVAIAGATV LALLPVIQTE RIDLQGNFED AIGLPLKLAEM
 SPGGQPWSHE ISAALGVLT IALIASVESL LSAVGVDKLH HGPRTDFNRE
 MVGQGSANVV SGLLGLLPIT GVIVRSSANV AAGARTRMST **ILHGVWILLF**
ASLFTNLVEL IPKAAALAGLL IVIGAQLVKL AHIKLAWRTG NFVYAITIV
CVVFLNLEGG YAIGLVVAIV FLVVRVVRAP VEVKPVGGEQ SKRWVRVDIDG
 TSLFLLPRL TTVLSKLPEG SEVTLNLDAD YIDDSVEAI SDWRAHETR
 GGVAIVETS PAKLHHAHAR PPKRHFASDP IGLVPPWRSAR GKDRGSASVL
 DRIDEYHRNG AAVLHPHIAG LTDSQDPYEL FLTCDARSIL PNVITASGPG
 DLYTVRNLGN LVPTDPDDRS VDAALDFAVN QLVGVSSVVVC GHSSCAAMTA
 LLEDDPANTT TPMMRWLENA HDLSLVYFRNH HPARRSAESA GYPEADQLSI
 VNVAQVERL TRHPILATAV AAADLQVIGI FFDISTARVY EVGPNGIICP
 DEPADRPVDH ESAQ

FIG. 6. An example of sequence coverage and TMHs of the membrane proteins identified. The underlined sequences indicate the identified peptides in this study. The yellow background shows where the transmembrane helices are located. Red lettering in the yellow background boxes indicates the end inside of the membrane. a, protein ctpV (Rv0969). b, hypothetical membrane protein Rv3273.

Single Peptide-based Identification—Because of the high specificity of peptide sequencing using the tandem MS technique, certain proteins could be identified by single peptide sequence matches when good quality MS/MS spectra were provided (23, 30, 31). There were an additional 237 proteins identified by matches with single peptide MS/MS spectra as listed in Supplementary Table II. Among them, there are 35 transmembrane proteins with various numbers of TMHs from 1 to 14.

DISCUSSION

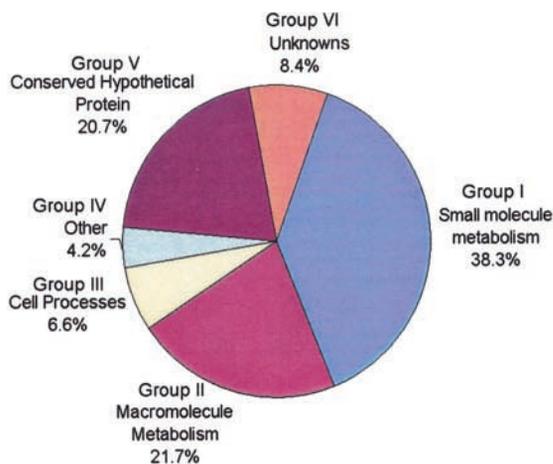
Although 2-DE provides high resolution separation of complex mixtures through both the pI and molecular mass dimensions, the efficiency of membrane protein identification has been low because of their hydrophobicity and low abundance (18, 32). Based on our experiences with the 2-DE-based approach, even though thiourea was used to increase protein solubility, only a few of the membrane or membrane-associated proteins were identified. The 1D SDS gel separates proteins with various physical properties for analysis without any pI discrimination. Our profile contained 66 proteins with basic pIs higher than 10, which usually are undetectable by 2-DE-

based approaches. The strong SDS detergent solubilizes hydrophobic membrane proteins and helps to keep them moving in the gel. Inevitably, a single 1D SDS band may contain several proteins that lead to a mixture of peptides in in-gel digestion. However, the microcapillary LC setup prior to nano-spray-MS/MS analysis provides the high resolution peptide separation and detection sensitivity required for the analysis of complex peptide mixtures. With the additional molecular mass information given by the gel, this method has proven to be as effective and efficient as the gel-free 2D LC method that had been successfully applied to complex protein mixtures (31) and membrane proteome (33) analysis.

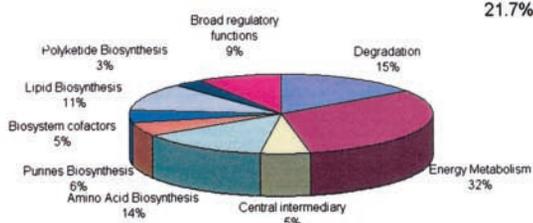
Of the 739 proteins identified unambiguously, more than 400 proteins were found for the first time in a proteomic study. Of the 85 transmembrane proteins identified, only six of them have been reported previously (Table II and Supplementary Table I). These results suggest that this 1D gel-LC-MS/MS approach provides a far more comprehensive membrane protein profile than 2-DE-based approaches. Of note is that 32 of the transmembrane proteins identified belong to the functional subgroup of the cell envelope (Group II.C, Fig. 7b). Of the 787 total genome-based predicted transmembrane proteins, 244 proteins fall into this functional category. Therefore, about 13.1% (32 of 244) of the total predicted proteins in this functional category were identified. In addition, if the proteins identified through the single peptide matches are included, then an additional 10 transmembrane proteins will be included in this functional group. Furthermore, it should be noted that although the membrane proteins represent only about 12% of the overall identified proteomic profile, many other proteins were found to be membrane-associated. For example, 90 proteins were identified as being involved in energy metabolism, and another 32 proteins could be functional in lipid biosynthesis. Because of their functional categories, it can be postulated that these proteins are intensively interacting with the MTB membrane and participating in membrane biochemical processes.

Because certain membrane proteins play vital roles in many cellular processes, they may become important targets for diagnosis and therapeutics. Some molecular approaches target receptors, such as ligand-gated ion channel and voltage-dependent ion channel proteins located in membranes (34). In our proteomic profile, multiple protein factors involved in these processes were simultaneously identified. For example, because of its high specificity, the PhoS antigen is widely used in the serodiagnosis of TB. However, because of the lack of sensitivity the detection of this antigen for diagnosis is not very reliable (35). Many studies have searched for novel sensitive, immunodominant MTB antigens. MTSA-10 (Rv3874) was found to be hypersensitive in TB-infected but not in vaccine-immunized animals (36). In another study directed toward improving the accuracy of serological diagnosis for TB, nine novel antigens including ggtB (Rv2394), lprF (Rv1368), lpqD (Rv3390), and Rv0875c were found to be

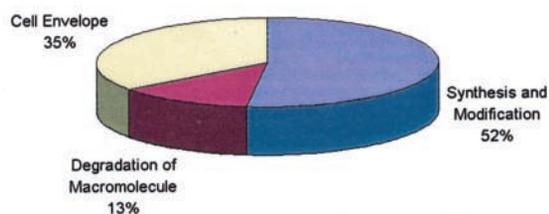
(a) Functional Category



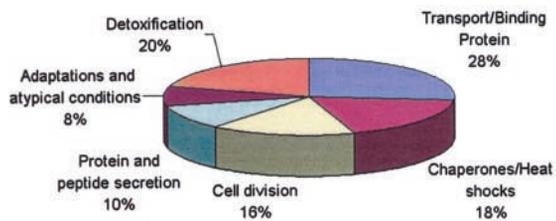
(b)



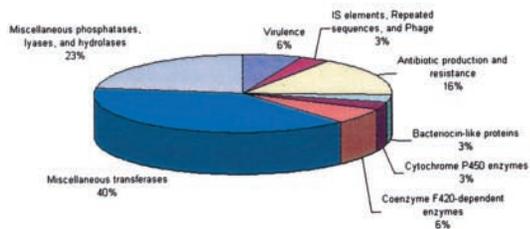
Major Functional Group I



Major Functional Group II

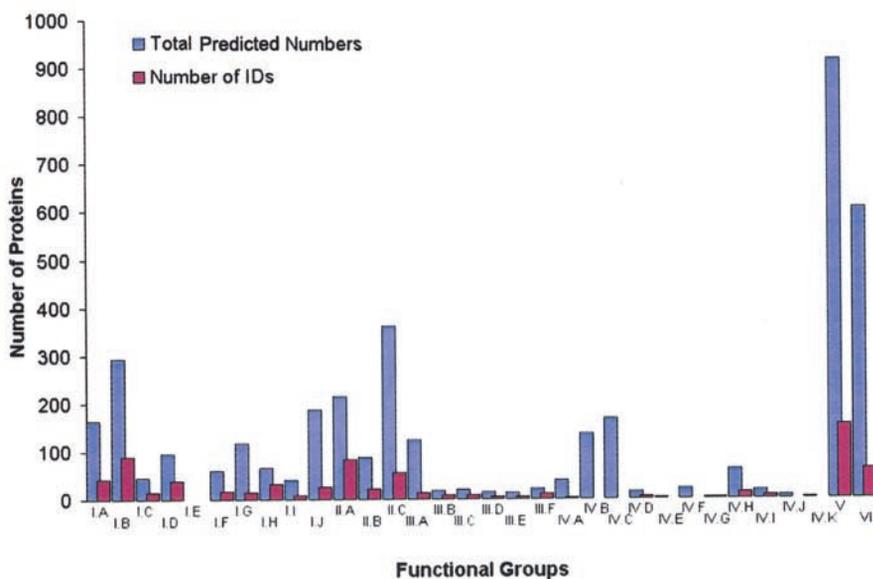


Major Functional Group III



Major Functional Group IV

(c)



potential candidates for a serodiagnostic test (37). The early secreted protein ESAT6 (Rv3875) that can stimulate CD4⁺ T-cells from tuberculosis patients was also identified in our study. Importantly, ESAT6 can also induce the mRNA transcripts of the macrophage inflammatory proteins, monocyte chemotactic protein and IFN- γ , in peripheral blood mononuclear cells from tuberculosis patients (38, 39). With proteomic validation, ESAT6 can be proposed to be an antigen for TB clinic detection. Antigen 85A (Rv3804c) can be recognized by antibodies from an HIV-infected TB patient (35). In a study of TB comparative gene expression, antigen 85A and rpoB (Rv0667) were found to be expressed inside macrophage cells, whereas antigen 85B (Rv1886c) and rpoV (Rv2703, sigA, RNA polymerase σ factor) were found to be expressed only in the *in vitro* culture (40). This type of study can help us to understand the factors involved in the MTB response to external stress. All the proteins mentioned above were identified in this study. In addition, these proteins, such as IprF, IpqD, Rv0875c, antigen 85A, and antigen 85B, all possess transmembrane helices. Thus, many important proteins with clinical potential involve cell membrane-associated processes.

CONCLUSION

For the first time, we have obtained a relatively comprehensive protein profile for the membrane composition of the MTB H37Rv strain. Including those proteins with single peptide matches, there are 976 unique proteins identified through the integrated 1D SDS-LC-MS/MS approach. Based on the 1D SDS-PAGE coupled with the high throughput μ LC-nanospray-MS/MS method, the complex proteome of the MTB membrane fraction was profiled, and novel proteins were identified. Proteins with serodiagnostic potential were identified that both transmembrane proteins and non-membrane proteins contained in membrane fractions. These novel proteins identified in our proteomic approach have provided new insights into the MTB membrane proteome and provide antigen, vaccine, and anti-TB drug targets.

Acknowledgment—We thank Dr. Kwasi G. Mawuenyega for help with data searches.

* This work was supported by Los Alamos National Laboratory Directed Research Development Grants 20030508ER and 20020048DR

and by United States Department of Energy Grants ERW9840 and KP1103010 (to X. C.). This is publication No. LA-UR-03-4877 from the Los Alamos National Laboratory. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

☐ The on-line version of this article (available at <http://www.mcponline.org>) contains Supplementary Tables I and II.

|| Recipient of a Presidential Early Career Award for Scientists and Engineers (PECASE, 2000–2005). To whom correspondence should be addressed: B-2, MS M888, Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87544. Tel.: 505-665-3197; Fax: 505-665-3024; E-mail: chen_xian@lanl.gov.

REFERENCES

1. World Health Organization (2003) *Global Tuberculosis Control: Surveillance, Planning, Financing (WHO Report 2003)*, Geneva
2. Russell, D. G. (2001) Mycobacterium tuberculosis: here today, and here tomorrow. *Nat. Rev. Mol. Cell Biol.* **2**, 569–577
3. Rhen, M., Eriksson, S., Clements, M., Bergstrom, S., and Normark, S. J. (2003) The basis of persistent bacterial infections. *Trends Microbiol.* **11**, 80–86
4. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., III, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., et al. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**, 537–544
5. Pandey, A., and Mann, M. (2000) Proteomics to study genes and genomes. *Nature* **405**, 837–846
6. Yoshida, M., Loo, J. A., and Lepley, R. A. (2001) Proteomics as a tool in the pharmaceutical drug design process. *Curr. Pharm. Des.* **7**, 291–310
7. Jungblut, P. R., Schaible, U. E., Mollenkopf, H. J., Zimny-Armdt, U., Raupach, B., Mattow, J., Halada, P., Lamer, S., Hagens, K., and Kaufmann, S. H. (1999) Comparative proteome analysis of Mycobacterium tuberculosis and Mycobacterium bovis BCG strains: towards functional genomics of microbial pathogens. *Mol. Microbiol.* **33**, 1103–1117
8. Rosenkrands, I., Weldingh, K., Jacobsen, S., Hansen, C. V., Florio, W., Gianetri, I., and Andersen, P. (2000) Mapping and identification of Mycobacterium tuberculosis proteins by two-dimensional gel electrophoresis, microsequencing and immunodetection. *Electrophoresis* **21**, 935–948
9. Rosenkrands, I., King, A., Weldingh, K., Moniatte, M., Moertz, E., and Andersen, P. (2000) Towards the proteome of Mycobacterium tuberculosis. *Electrophoresis* **21**, 3740–3756
10. Covert, B. A., Spencer, J. S., Orme, I. M., and Belisle, J. T. (2001) The application of proteomics in defining the T cell antigens of Mycobacterium tuberculosis. *Proteomics* **1**, 574–586
11. Mattow, J., Jungblut, P. R., Schaible, U. E., Mollenkopf, H. J., Lamer, S., Zimny-Armdt, U., Hagens, K., Muller, E. C., and Kaufmann, S. H. (2001) Identification of proteins from Mycobacterium tuberculosis missing in attenuated Mycobacterium bovis BCG strains. *Electrophoresis* **22**, 2936–2946
12. Mattow, J., Jungblut, P. R., Muller, E. C., and Kaufmann, S. H. (2001) Identification of acidic, low molecular mass proteins of Mycobacterium

FIG. 7. **The functional category distribution of the 739 identified proteins.** Assignments were made based on the Sanger Institute gene database. The distributions are among (a) the major functional groups and (b) the subgroups within each major functional group (Groups I–IV). The percentage for each subgroup indicates the percentage of the total number of identified proteins in its major functional group. c, comparison of the total predicted protein number based on the genome data and the number of proteins identified in each sub-functional group. *I.A*, small molecule degradation; *I.B*, energy metabolism; *I.C*, central intermediary metabolism; *I.D*, amino acid biosynthesis; *I.E*, polyamine synthesis; *I.F*, purines, pyrimidines, nucleosides, and nucleotides; *I.G*, biosynthesis of cofactors, prosthetic groups, and carriers; *I.H*, lipid biosynthesis; *I.I*, polyketide and non-ribosomal peptide synthesis; *I.J*, broad regulatory functions. *II.A*, synthesis and modification of macromolecules; *II.B*, degradation of macromolecules; *II.C*, cell envelope. *III.A*, transport/binding proteins; *III.B*, chaperones/heat shock; *III.C*, cell division; *III.D*, protein and peptide secretion; *III.E*, adaptations and atypical conditions; *III.F*, detoxification. *IV.A*, virulence; *IV.B*, IS elements, repeated sequences, and phage; *IV.C*, PE and PPE families; *IV.D*, antibiotic production and resistance; *IV.E*, bacteriocin-like proteins; *IV.F*, cytochrome P450 enzymes; *IV.G*, coenzyme F420-dependent enzymes; *IV.H*, miscellaneous transferases; *IV.I*, miscellaneous phosphatases, lyases, and hydrolases; *IV.J*, cyclases; *IV.K*, chelatases. *V*, conserved hypothetical proteins; *VI*, unknowns.

- tuberculosis strain H37Rv by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry. *Proteomics* **1**, 494–507
13. Jungblut, P. R., Muller, E. C., Mattow, J., and Kaufmann, S. H. (2001) Proteomics reveals open reading frames in *Mycobacterium tuberculosis* H37Rv not predicted by genomics. *Infect. Immun.* **69**, 5905–5907
 14. Sigler, K., and Hofer, M. (1997) Biotechnological aspects of membrane function. *Crit. Rev. Biotechnol.* **17**, 69–86
 15. Camacho, L. R., Ensergueix, D., Perez, E., Gicquel, B., and Guilhot, C. (1999) Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. *Mol. Microbiol.* **34**, 257–267
 16. Daffe, M., and Etienne, G. (1999) The capsule of *Mycobacterium tuberculosis* and its implications for pathogenicity. *Tuber. Lung Dis.* **79**, 153–169
 17. Sinha, S., Arora, S., Namane, A., Pym, A. S., and Cole, S. T. (2002) Proteome analysis of the plasma membrane of *Mycobacterium tuberculosis*. *Comp. Funct. Genomics* **3**, 470–483
 18. Santoni, V., Molloy, M., and Rabilloud, T. (2000) Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**, 1054–1070
 19. Simpson, R. J., Connolly, L. M., Eddes, J. S., Pereira, J. J., Moritz, R. L., and Reid, G. E. (2000) Proteomic analysis of the human colon carcinoma cell line (LIM 1215): development of a membrane protein database. *Electrophoresis* **21**, 1707–1732
 20. Pan, S., Gu, S., Bradbury, E. M., and Chen, X. (2003) Single peptide-based protein identification in human proteome through MALDI-TOF MS coupled with amino acids coded mass tagging. *Anal. Chem.* **75**, 1316–1324
 21. Mortz, E., Krogh, T. N., Vorum, H., and Gorg, A. (2001) Improved silver staining protocols for high sensitivity protein identification using matrix-assisted laser desorption/ionization-time of flight analysis. *Proteomics* **1**, 1359–1363
 22. Hunter, T. C., Yang, L., Zhu, H., Majidi, V., Bradbury, E. M., and Chen, X. (2001) Peptide mass mapping constrained with stable isotope-tagged peptides for identification of protein mixtures. *Anal. Chem.* **73**, 4891–4902
 23. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., and Gygi, S. P. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* **2**, 43–50
 24. Schwartz, R., Ting, C. S., and King, J. (2001) Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. *Genome Res.* **11**, 703–709
 25. Chen, C. P., and Rost, B. (2002) State-of-the-art in membrane protein prediction. *Appl. Bioinformatics* **1**, 21–35
 26. Moller, S., Croning, M. D., and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics (Oxf.)* **17**, 646–653
 27. Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182
 28. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580
 29. Graham, J. E., and Clark-Curtiss, J. E. (1999) Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proc. Natl. Acad. Sci. U. S. A.* **96**, 11554–11559
 30. Shevchenko, A., Loboda, A., Ens, W., and Standing, K. G. (2000) MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. *Anal. Chem.* **72**, 2132–2141
 31. Washburn, M. P., Wolters, D., and Yates, J. R., III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–247
 32. Herbert, B. R., Harry, J. L., Packer, N. H., Gooley, A. A., Pedersen, S. K., and Williams, K. L. (2001) What place for polyacrylamide in proteomics? *Trends Biotechnol.* **19**, S3–S9
 33. Wu, C. C., MacCoss, M. J., Howell, K. E., and Yates, J. R. (2003) A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotechnol.* **21**, 532–538
 34. Herz, J. M., Thomsen, W. J., and Yarbrough, G. G. (1997) Molecular approaches to receptors as targets for drug discovery. *J. Recept. Signal Transduct. Res.* **17**, 671–776
 35. Samanich, K. M., Keen, M. A., Vissa, V. D., Harder, J. D., Spencer, J. S., Belisle, J. T., Zolla-Pazner, S., and Laal, S. (2000) Serodiagnostic potential of culture filtrate antigens of *Mycobacterium tuberculosis*. *Clin. Diagn. Lab. Immunol.* **7**, 662–668
 36. Colangeli, R., Spencer, J. S., Bifani, P., Williams, A., Lyashchenko, K., Keen, M. A., Hill, P. J., Belisle, J., and Gennaro, M. L. (2000) MTSA-10, the product of the Rv3874 gene of *Mycobacterium tuberculosis*, elicits tuberculosis-specific, delayed-type hypersensitivity in guinea pigs. *Infect. Immun.* **68**, 990–993
 37. Moran, A. J., Treit, J. D., Whitney, J. L., Abomoelak, B., Houghton, R., Skeiky, Y. A., Sampaio, D. P., Badaro, R., and Nano, F. E. (2001) Assessment of the serodiagnostic potential of nine novel proteins from *Mycobacterium tuberculosis*. *FEMS Microbiol. Lett.* **198**, 31–36
 38. Ulrichs, T., Munk, M. E., Mollenkopf, H., Behr-Perst, S., Colangeli, R., Gennaro, M. L., and Kaufmann, S. H. (1998) Differential T cell responses to *Mycobacterium tuberculosis* ESAT6 in tuberculosis patients and healthy donors. *Eur. J. Immunol.* **28**, 3949–3958
 39. Ulrichs, T., Anding, P., Porcelli, S., Kaufmann, S. H., and Munk, M. E. (2000) Increased numbers of ESAT-6- and purified protein derivative-specific gamma interferon-producing cells in subclinical and active tuberculosis infection. *Infect. Immun.* **68**, 6073–6076
 40. Mariani, F., Cappelli, G., Riccardi, G., and Colizzi, V. (2000) *Mycobacterium tuberculosis* H37Rv comparative gene-expression analysis in synthetic medium and human macrophage. *Gene (Amst.)* **253**, 281–291